# Infosys Internship 5.0

## **Title: Project Documentation: Air Quality Index Visualization**

Name : YASWANTH KUMAR K

**Introduction**

The *Air Quality Index Dashboard* project provides a visual representation of air quality data, focusing on trends, pollutants, and geographic variations. With increasing concerns about pollution's impact on health and the environment, this project empowers users to explore and understand AQI patterns through an interactive dashboard.

Designed using Power BI, the dashboard transforms complex datasets into user-friendly visualizations, making the data accessible for policymakers, researchers, and general users. By offering insights into pollutant trends and city-wise AQI comparisons, this project aims to support informed decision-making.

The development process included:

* Collecting and cleaning historical AQI data.
* Designing an intuitive and dynamic dashboard.
* Ensuring functionality and usability for diverse audiences.

This project was completed as part of the Infosys Springboard internship, showcasing skills in data visualization, analytical thinking, and technical development.

**Project Scope:**

Project Overview:

This project focuses on analyzing AQI data to identify regional and temporal patterns of air pollution. It provides an accessible platform for exploring how air quality changes over time and the relative contributions of pollutants.

Inclusions:

* An interactive dashboard displaying:
  + AQI trends over time.
  + Contributions of pollutants (e.g., PM2.5, NO₂, SO₂).
  + City-wise comparisons of AQI levels.
* Filters to explore data by city, year, and pollutant.
* Visualizations like bar charts, line graphs, and heatmaps.

Exclusions:

* Real-time AQI updates.
* Global data or predictive analysis.

Challenges

* Limited data availability for some pollutants.
* Handling large datasets while maintaining dashboard responsiveness**.**

**Requirements:**

Functional Requirements:

1. An interactive Power BI dashboard to display AQI data.
2. Filters for region, time period, and specific pollutants.
3. Downloadable reports for offline analysis.

Non-Functional Requirements:

1. User-friendly and intuitive design for easy navigation.
2. Quick load times for visualizations and filters.
3. Compatibility with modern browsers and devices.

User Stories:

* *As an environmental analyst*, I want to compare AQI levels across cities to prioritize resources.
* *As a policymaker*, I want to identify pollution trends over time to make data-driven decisions.
* *As a general user*, I want to learn about air quality in my area.

**Technical Stack:**

* Programming Languages: Python for preprocessing.
* Visualization Tool: Power BI for dashboard creation.
* Libraries: Pandas for data manipulation.
* Datasets: cleaned\_air\_pollutiondata.xlsx and air-flow-index.csv.
* Environment: Power BI Desktop for design and Power BI Service for sharing.

**Architecture:**

The project workflow follows these steps:

1. **Data Collection**: Historical AQI data was sourced from reliable datasets.
2. **Data Cleaning and Preprocessing**: Python was used to clean missing values, remove duplicates, and structure the data.
3. **Visualization**: Power BI was utilized to design an interactive dashboard with various charts and filters.
4. **Sharing**: The dashboard was published and shared with stakeholders via Power BI Service.

### 

**Development:**

Technologies Used:

* Python: For preprocessing AQI data, handling missing values, and generating clean datasets.
* Power BI: For creating dynamic charts and interactive filters.

Challenges and Solutions:

1. Data Quality Issues: Missing values and inconsistent formats.
   * *Solution*: Used Python to fill gaps and standardize the dataset.
2. Design Complexity: Balancing detailed visuals with simplicity.
   * *Solution*: Iterated on designs to enhance readability and functionality.

Best Practices:

* Ensured consistency in data labels and formats.
* Applied clear and concise labeling in charts for better user experience.

**Testing:**

Testing Approach:

1. Data Validation: Verified accuracy after preprocessing.
2. Dashboard Testing: Assessed the responsiveness of filters and interactivity.
3. Performance Testing: Measured loading times for larger datasets.

Results:

* All functionalities, including filters and charts, worked as intended.
* Data accuracy was verified against original datasets.

**Deployment:**

Process:

1. The Power BI dashboard was published to Power BI Service.
2. Access was shared via a secure link.
3. Final testing was conducted to ensure accessibility and functionality.

Instructions:

* Open the shared dashboard link.
* Use the filters to explore AQI trends by year, pollutant, and region.

**User Guide:**

1. Accessing the Dashboard:
   * Open the provided link in any modern browser.
2. Features:
   * Select a region or pollutant to explore specific data.
   * Hover over charts for detailed values.
3. Troubleshooting:
   * Refresh the page if charts do not load correctly.

**Conclusion:**

This project provides a robust and interactive platform for exploring AQI data. By offering dynamic visualizations and actionable insights, it empowers users to make informed decisions about air quality management.

Achievements:

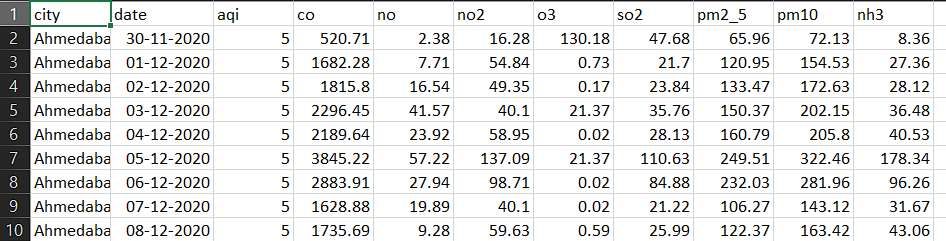
* Delivered a fully functional Power BI dashboard.
* Enabled in-depth exploration of air quality trends.

Future Enhancements:

* Integrate real-time data using APIs.
* Expand the scope to include global air quality data.

### **Appendices:**

**Sample Data Snippet :**

****

**Code Snippet for Data Cleaning:**

# Python script for handling missing data in AQI dataset

import pandas as pd

# Load dataset

data = pd.read\_csv('air\_quality\_data.csv')

# Handle missing values

data.fillna(method='ffill', inplace=True)

# Filter relevant columns

cleaned\_data = data[['Location', 'Year', 'AQI\_Value', 'PM2.5', 'PM10', 'SO2']]

cleaned\_data.to\_csv('cleaned\_aqi\_data.csv', index=False)

**References:**

* **Dataset Sources**:
  + CentralPollution Control Board (CPCB), India: <https://www.cpcworldwide.com/>
  + OpenAQ Platform: <https://openaq.org/>
* **Tools**:
  + Power BI Documentation: <https://learn.microsoft.com/en-us/power-bi/>
  + Python Pandas Library: https://pandas.pydata.org/